The Network Oracle

Joseph M. Hellerstein* † Vern Paxson‡ Larry Peterson§ Timothy Roscoe†
Scott Shenker* ‡ David Wetherall¶

*UC Berkeley †Intel Research, Berkeley ‡ICSI §Princeton University ¶University of Washington

Abstract

This paper sets out a high-level research agenda aimed at building a collaborative, global end-system monitoring and information infrastructure for the Internet's core state. We argue that such a system is beneficial, feasible and timely, representing an important area for engagement across database and networking technologies. We start by hypothesizing the benefits to the Internet of a "Network Oracle" that could answer real-time questions about the global state of the network. We then argue that database and networking technology should make it possible to provide a useful approximation of such an oracle today, gathering information from a large number of end hosts and delivering useful views to each end system. We further argue that this can and should be done in a decentralized fashion. We provide an outline of such a system: it employs sensing agents, along with a distributed query/trigger/dissemination engine, and possible attractive end-user applications. A key point of our discussion is the timeliness and importance of a grassroots agenda for Internet monitoring and state-sharing. While significant social and economic barriers to deploying a centralized, public Internet monitoring infrastructure exist, there are corresponding social and economic incentives for a collaborative approach.

1 Vision

There has been increasing interest in recent years in topics at the nexus of distributed databases and core networking, and various agendas have been laid out for exploring synergies ([18, 11, 23, 13]). To date, however, there has been little work on rethinking the Internet architecture in response to these technical directions.

In this paper, we lay out one such agenda in broad terms. We propose that the database, networking, and distributed systems research communities collaborate in building a global end-system monitoring and information infrastructure for the Internet's core state.

Our discussion begins with a thought experiment. Setting aside concerns about social and technical barriers, suppose there existed a Network Oracle: a queryable object that any end-system on the Internet could use to immediately receive information about global network state, from the recent past to real-time updates. This state could include complete network maps (including addressing realms and NAT gateways), link loading, point-to-point latency and bandwidth measurements, event detections (e.g., from firewalls), naming (DNS, ASes, etc.), end-system software configuration information, even router configurations and routing tables.

This is considerably more information than is available to end-systems today. The existence of the Network Oracle would allow end-systems to make more sophisticated decisions about every aspect of their interaction

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

with the network: the parties they communicate with, the routes and resources they use, and the qualities of the various actors in the communication chain. In Section 2 we give concrete examples of how end-systems and end-users could benefit from this information.

How far away from this vision are we today? Network monitoring is not a new activity. Many parties collect significant information in today's Internet, including carriers and large IT departments. However, the information collected is by no means comprehensive; it is chosen with relatively narrow goals in mind, usually with a focus on backbone traffic engineering and academic networking research. Also, since the data is typically collected "in the middle" of the network, it only captures packets as they traverse those links; it misses significant information about the properties and traffic in small Intranets, in switched subnets of large Intranets, in WiFi communities, and in similar rich and evolving "microclimates" at the edges of today's Internet.

Furthermore, current network monitoring systems focus on data collection, but ignore public-access query or dissemination facilities. Information gathered by today's network monitors is generally available neither to end-users nor their protocols or applications. This places inherent limits on innovation. Making this information widely available in near-real-time can significantly change the protocol and distributed system design landscape, in a way that offline centralized analysis cannot. Today's Internet was designed under the assumption that it is not feasible to gather and disseminate such information at scale, and researchers and developers of end-user network applications constrain their design space accordingly. We argue below that this assumption no longer holds. Eliminating these constraints can open up new opportunities for significant innovation in network functionality and robustness – opportunities that span traditional boundaries between data management, networking, and distributed systems.

2 What For?

We conjecture that a Network Oracle would have multiple benefits for various parties at differing timescales. These include (a) social benefits such as raised awareness of security risks and the importance of end-system maintenance, (b) medium-term network engineering benefits including localization of performance problems and identification of malevolent actors, and (c) long-term design opportunities including the development of end-to-end network protocols and distributed applications that leverage the global information. In this section we present several examples.

Performance Fault Diagnosis. There are very few options for diagnosing performance faults within the Internet today. When a site loses connectivity, it is possible to use tools such as traceroute to determine whether the problem is likely local or remote. But when the Internet performs poorly from a given site it is difficult to determine whether the problem lies with that site or elsewhere, and whether there is in fact a problem to be corrected rather than a temporary overload. The key difficulty is that there is often no baseline that the site can use to gauge what level of performance it should obtain.

The Network Oracle we envision addresses this difficulty directly by allowing a site to compare and contrast its performance with that of other sites, both in the past and at the given moment in time. For instance, a site might note suspect destinations for which it obtains persistently low performance, and query the oracle for these suspect destinations. If these destinations subscribe to the oracle, then current and historical information will be available to assess the overall level of performance of the destination and whether it matches past performance. Recent falloffs suggest a problem with the site; consistent performance suggests a problem elsewhere, and if all other sites reporting information about the suspect destination are receiving equally poor performance, then the problem likely lies with the site itself which is simply overloaded. This same test can be applied to regions of the network between the site and destinations. The effect is to narrow the region of performance faults, facilitating correction, as well as to identify alternative paths that offer improved performance. It is the sharing of performance data across sites facilitated by the oracle that makes this possible.

Tracking attacks. Hosts on the Internet experience incessant attack [20]. This engenders an edginess in

Internet users, much of which boils down to questions like: *(i)* Is this remote host attempting to contact me a bad guy? *(ii)* Am I being targeted, or just enduring what everyone else endures? *(iii)* Is there a new type of attack going on? *(iv)* Is something happening on a global Internet scale?

By providing insight into activity experienced by one's peers and Internet sites in general, the Network Oracle can help answer questions both about types of activity, and which hosts have been seen doing what. This kind of shared information is not only interesting, but potentially actionable. As a simple example, studies have demonstrated that very few source IPs generate a significant fraction of port scans [32]. A real-time distributed query could compute the global "worst offenders" list, and allow firewalls at the endpoints to adaptively change their packet filters to track changes in the list. Placing the Network Oracle at the core of a global security agenda raises major research issues concerning attacks on and subversion of the infrastructure and its data, and in general ensuring that the information has an "actionable" degree of fidelity. We return to this point in Section 5.

Network Routing Protocols. The current routing infrastructure computes routing tables that, to a first approximation, are applied to all flows headed to the same destination. However, it is clear that no uniformly applied routing protocol, no matter how well designed, can accommodate every flow's policy and/or performance requirements. Starting with source routing, and continuing with more recent designs such as NIRA [31] and TRIAD [12], there is a large literature about mechanisms that would allow flows to choose their own route. Research has typically focused on mechanisms for expressing and implementing the desired route, but much less attention has been paid to how flows (or the hosts acting on their behalf) could determine which route would best serve their needs. If only a very small fraction of flows were making such individualized route choices, then fairly primitive, and bandwidth-expensive, route exploration mechanisms could be used. But if source-specified routing became commonplace, then a more scalable approach would be needed. In particular, one would want the information used to decide routes to be shared, rather then individually discovered. The Network Oracle approach, in which general classes of information are gathered and made available, could provide a virtual repository for the relevant information. Initial work (e.g., [16]) suggests that such a repository, along with a general querying facility, could be used for such route computation.

Adaptive Applications. A natural extension of having collected a wealth of information about the health of the network is for applications to adapt; to react to this information by selecting alternative protocols, alternative routes, or even alternative sources for the content they are trying to access. While adaptation might happen purely at the end system (e.g., selecting a variant of TCP most suitable for the current end-to-end path), it is easy to imagine the emergence of way-stations that help end systems avoid network trouble spots and rapidly recover from failure [25, 1], as well as more globally coordinated network services that are able to distribute network load over a wider swath of the Internet [19, 9].

An Internet Screensaver. Distributed computation projects like SETI@Home [2] have demonstrated that individuals will contribute private computing resources to the common good – particularly if rewarded with the right combination of entertainment (e.g., interesting screensavers) and community-building tools (e.g., the SETI "leader board" listing the top contributors). Given mounting press and public concern about Internet viruses and worms, the time seems ripe to build an Internet Screensaver – a peer-to-peer application of end-hosts monitoring the network for security events and performance anomalies. Such an application could have multiple benefits. First, it could serve as an attractive, sizable testbed for a prototype Network Oracle, measuring the network from the richness of a variety of "last-mile" endpoints. Second, if properly designed it could engender a unique culture of enlightened vigilance, with client machines swapping notes on anomalous traffic for a variety of purposes. For example, end-users could set up social networks for "community watch", actively probing each other's machines for vulnerabilities. They could swap notes on passively-monitored undesirable traffic (worms, port-scans, spam), to help configure firewall rules. They could compare performance across ISPs. While they may not provide the most accurate measurements or the most effective security measures, these techniques would give Internet users insight into their own experience and incentive to control it more carefully.

Serendipity. While the previous scenarios described practical uses for the Network Oracle, its relevance to networking and database research should not be neglected. The network measurement literature is huge, and continues to grow at an astounding rate, so the field is hardly lacking for interesting questions to ask and relevant data to answer them. Database technologies like streaming query systems are finding some of their most compelling applications in these contexts. However, much of the networking research analyzes data that was gathered with the specific question in mind, or at least in a specific context with limited scope; for example, a routing study might collect BGP routing tables but would be unlikely to also simultaneously collect data from firewalls or application logs at points nearby the relevant routers. Thus, current measurement studies have a naturally limited ken which may prevent certain networking questions from being answered, or even being asked. This presents a chicken-and-egg problem with respect to database research ideas as well; the lack of Internet-scale distributed database products prevents network researchers from asking these kinds of questions, which in turn leaves (some) database researchers and developers unmotivated to design systems for this purpose. The result is that a ”virtuous cycle” of collaboration between the communities has been relatively slow to emerge. We hope and anticipate that, should it be built, a general-purpose Network Oracle could open up surprising new connections – both for research and for application.

3 Why Now?

The vision of making detailed, relevant information about global network state available at all endpoints was not realistic several years ago. We argue that the vision is now realistic, and is becoming easier.

The Social Case. Our first point is social, not technical. A widespread monitoring infrastructure will only be achieved (and sustained) if the sensors are widely desired and deployed. We believe that the time is ripe for encouraging end-users to install software to help improve the Internet.

We are at a point in Internet history where the need for protective action has hit the common consciousness. Spam, viruses and worms have led even unsophisticated users to take non-trivial technical steps to try and ameliorate these problems. Simultaneously, news stories about identity theft, Internet credit card fraud and the U.S. Patriot Act have raised popular sensitivity to the importance of the Internet in their lives. We believe that many users are not only willing and ready, but thirsty to install applications that can improve their trust in the Internet, and better inform them of risks.

Working against this, of course, is concern with individual privacy – often argued to be in tension with safety. A recent article [4] suggests that the strategic need to secure the Internet, combined with the ease of surveillance as a tool for doing so, will lead to an Internet that is a “broadly surveilled police state”. Is it possible for the community to be vigilant without compromising their privacy?

This can be construed as a technical or social challenge; ultimately it is both. The technical agenda in privacy-preserving information sharing is heating up (see the proceedings of any database conference) but is in its infancy, and it may be years before practical guarantees can be made about the privacy of information in a Network Oracle. We conjecture that the deployment of an early-stage Network Oracle need not wait for such technology to mature. Many users may be willing to opt into a decentralized Network Oracle, even if they would not do so with a centralized system offered by a large organization (commercial or non-profit), based on the user’s presumption that no single malicious individual would see much of their information. This soft attitude toward privacy is surprisingly widespread – witness the number of people who identify their IP address to unknown peers in order to acquire copyrighted material illegally, or the often vigorous participation in pseudo-anonymous fora such as newsgroups and chat rooms. The social issue of privacy is less contentious when large interests (companies, governments, standards bodies) are removed from the debate. We explore the implications of the requirement for decentralization below.

Technology Trends. Our second point concerns the technical feasibility of disseminating adequate information about the state of the Internet. We believe that technology trends are working in our favor.

We conjecture that *the "metadata" of the Internet's behavior is shrinking relative to data being shipped across the Internet.* The bandwidth required to ship useful measurement data around the network to end users, as a fraction of the total bandwidth available to end users, is decreasing. This makes it more attractive to start making measurement data available to all end systems.

We note that while both bandwidth and flow size are increasing, the data required to describe a flow is not. Available network bandwidth is increasing, both in residential (DSL, Cable Modems) and business settings (100–1000 Mbps Ethernet), a trend even more pronounced in technologically advanced countries like South Korea and Japan. At the same time, the size of data objects transferred through these larger pipes is also increasing (video streams, large downloads, VoIP sessions, even web-page objects are becoming larger), while the amount of data required to describe this activity is staying relatively constant – it scales with the number of end systems, rather than the available link capacity. While there are counter-examples (such as the proliferation of instant-message traffic), in general the principle holds: the amount of data one needs to know at endpoints to make informed decisions is growing more slowly than the access bandwidth at these end systems.

Moreover, processing power in end systems (indeed, systems anywhere in the network) is outstripping the increase in available network bandwidth. We are a long way from the days when 80% of the CPU cycles of an Alto were devoted to running a 3Mb/s Ethernet interface. A modern IA32 server can comfortably handle a 1Gb/s Ethernet interface at line rate with most of its CPU cycles to spare. The conclusion is that systems can afford to give much more consideration to traffic (in particular, its temporal characteristics) than is typically assumed in protocols and implementations. A corollary is that users with more cycles to throw at the problem may be capable of extracting more value (per b/s) from their network link.

Technology Innovations. Our final point is that we are seeing the appearance of technologies that can take this opportunity and make it useful. In recent years the building blocks for delivering a modest but non-trivial approximation of the Network Oracle are falling into place, with contributions from a number of research communities: for example, content-addressable networks, distributed query processing, statistical data reduction techniques, and statistical machine learning techniques. We return to these in Section 5.

4 Why Us?

The research community is uniquely positioned to realize an initial approximation to the Network Oracle. Other parties have no incentive to pursue this agenda: while the long-term benefits are significant, neither network providers nor equipment vendors gain obvious advantage from investing in such an endeavor at this stage. By contrast, this is an opportunity for researchers: while the Network Oracle vision represents a shift in emphasis for some networking research groups, substantial new research agendas and synergies exist in this direction.

Historically, ISPs have been resistant to the sharing of measurement information, except as marketing and sales aids. Where they have instrumented their networks, it has been with the internal goal of traffic engineering. Carriers have little interest (for sound commercial reasons) in disseminating end-to-end performance or security measurements. Consequently, equipment vendors have little interest in the problem; indeed, a shared, global management infrastructure may be a disruptive technology that threatens their market position.

In short, we are in a situation where commercial benefits of change are indirect but communally valuable, however the commercial threats are direct. The Network Oracle is not going to happen commercially at first. However, the situation is very different for the networking research community. In particular, the fields of network measurement and security have much to gain from realizing a Network Oracle.

Internet measurement in academia has been heavily restricted by the forms of measurement to which it has access, with rising security and privacy concerns making this increasingly difficult rather than the situation easing over time. This often leads to work that is implicitly driven and shaped by (shrinking) measurement opportunities. Internet security research, on the other hand, has struggled with the rise of rapid, automated attack technology, the loss of a defensible "perimeter" with mobile Internet devices, and an inability to adequately track and share information about miscreants. Both communities stand to gain by realizing a Network Oracle.

The Network Oracle would act as a rallying point where unnecessarily divergent research thrusts can be brought back together: network measurement, security, distributed systems, distributed databases, and statistical methods. Network measurement researchers would gain access to much larger, shared datasets than available to them today. Researchers in data management and analysis would achieve renewed relevance by taking their technologies out of endpoint data systems and rethinking them at Internet-scale. Similarly, security researchers would gain the ability to directly tackle problems at global scale, and with global resources.

Research in turn performs a bootstrapping function. The Network Oracle puts monitoring, diagnosis, and measurement functionality directly into end systems. If it can demonstrate value to end users in this way, it provides a path by which many measurement, monitoring, and diagnostic techniques can achieve a critical deployment mass without first requiring productization. Organizations with strategic interests in the deployment of such techniques can thus derive immense benefit from "plugging into the information substrate." Products follow deployment, rather than the other way around.

5 How?

The Network Oracle will not appear overnight. Here we highlight some challenges in realizing this vision.

Adoption and Uptake. The Network Oracle is characteristically a shared infrastructure. For reasons we have already outlined, we expect, at best, limited data from ISPs to start with. Consequently, the Network Oracle will rely at first on end-systems for both data sources and applications (consumers of data).

Projects like NETI@home [24] and DIMES [26] are exploring large-scale network measurement from end hosts, and the DShield project [8] warehouses firewall data sent in from many sources. Bundling a sensor with a global query visualization like an Internet Screensaver seems like a good incentive here, with a quid-pro-quo opt-in model: for the features you publish, you can see distributed results involving such features.

Moreover, another rich source of endpoints is "dark" IP address space, as monitored by Network Telescopes. While the traffic at these addresses is idiosyncratic, it is of interest to many parties, and could serve as "seed" data to populate the screensavers of early adopters.

The Network Oracle must also be able to locate and interface with large curated databases of information as well as distributed real-time sources. This includes slowly-changing network data (e.g., WHOIS) and archival data warehouses of traffic information (e.g., RouteViews).

Finally, we note that the definition of end systems expands to include large distributed services like P2P networks and CDNs. Public-minded instances of these services can share interesting traffic data with the Network Oracle, along the lines of PlanetSeer [33].

Architecture and Deployment. We envision a healthy diversity of popular sensors targeted at different network features, sharing an integrated query processing backplane. The function of this backplane is the processing (filtering, summarization, correlation) of data from many sources, and the delivery of relevant results to interested end systems. The amount of source data involved means that computation needs to occur "in the network", along the network path from the data source to the consumer.

The deployment of this computational infrastructure could evolve in a number of ways. We do not foresee a centrally administered solution succeeding. One option is to have a consortium manage a well-provisioned infrastructure, conceivably federated in nature like the Internet today. The recent success of PlanetLab is encouraging in this regard, but it remains unclear whether a consortium can maintain a production service like the Network Oracle without sustainable, measurable benefit to the institutions hosting the machines.

An alternative is an organic p2p deployment, with the query processor being bundled with the sensors. This is easier to deploy than the "distributed platform" approach, and is in an important sense more self-sustaining: the system remains up as long as sensors are deployed. It has a populist flavor that may allay some concerns about privacy and control. Of course, robustness in the p2p approach raises many technical challenges of scale, management, and resistance to attack or manipulation.

Technical Approach. Challenging as the Network Oracle may seem from a technical standpoint, we think most of the pieces of the puzzle have fallen into place in recent years, and a focused research effort could bring together a usable system in the spirit of a Network Oracle in short order. We present a brief selection here.

First, distributed query processing and content-addressable networks (DHTs) are getting much better at providing the right information to the right place at the right cost. P2p systems like PIER [14] push query processing into the network to reduce the data shipped during query answering. In addition to queries, triggering functionality akin to active databases is also possible, even with conditions that span distributed data [15]. A key tenet is the *data independence* that underpins relational databases: the physical organization (e.g., network location) of data should be separate from the logical data model and query interface [13]. Queries can be posed on data regardless of its location, and data can be reorganized without requiring changes in queries or applications that embed them.

DHTs are the first technology that provide this kind of data independence at Internet scale. PIER's "flat" DHT infrastructure is potentially a better fit for the Network Oracle than hierarchically organized systems; it does not depend on a small number of "roots" for the information and processing as DNS does, nor does it restrict the system to queries that traverse the hierarchy (as does, e.g., Astrolabe [28] or IrisNet [7]).

Second, the practical application of statistical methods in systems has been maturing over the last decade, in particular in computing approximate answers to queries (e.g. [3, 10]). Early exploration of these techniques in distributed settings are promising (e.g., [17, 6]). Also, distributed implementations of techniques like graphical models are emerging in the machine learning community, largely in the sensor network space (e.g., [21]). These approaches model statistical correlations in data, and use the models to predict data values and quantify uncertainty; this is useful both for predicting missing data, and for "cleansing" noisy acquired data.

Third, the security and trustworthiness of the Network Oracle implies several key challenges: validating the fidelity of data and computations, managing resource consumption at the end-hosts, providing accountability of misbehaving components, ensuring that the system itself is not used as an attack platform, and a viable and enforceable privacy framework. Security in p2p systems has been a topic of interest in recent years, with progress being made on topics including self-certifying data, secure routing, and fair sharing of work (e.g. [29]).

Finally, the research community now has the resources to do non-trivial test deployments of global-scale systems in controlled, repeatable "laboratory" settings [30, 27] and more permanently in the real Internet [22].

We view the development and deployment of a limited-function but useful Network Oracle as more actionable – and hence more fertile – than the recent proposal for a "knowledge plane" [5]. That vision is rather more broad, but also more vague. We favor an approach centered on workable designs and working implementations, which can spur organic community collaboration and follow-on work.

6 Conclusion

In speculating about a Network Oracle, our goal is not to map out an unattainable research ideal. On the contrary, a useful approximation of the vision is broadly desirable, and eminently feasible.

A research agenda for the community in this direction must encompass a range of areas, including overlay routing, query processing and database management, network measurement techniques, distributed intrusion detection, distributed statistical methods, and a broad set of issues in security and privacy. To achieve impact, this research must be informed and complemented by a thoughtful strategy for gaining and sustaining uptake in the system infrastructure. Such an effort would benefit significantly from the research community rallying around the problem, seeding the space with interesting data streams and useful compute resources, and working together toward common techniques and protocols.

References

[1] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris. Resilient Overlay Networks. In *ACM SOSP*, 2001.

[2] D. P. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer. SETI@home: An Experiment in Public-Resource Computing. *CACM*, 45(11):56–61, November 2002.

[3] D. Barbará, *et al.* The New Jersey Data Reduction Report. *IEEE Data Eng. Bull.*, 20(3), 1997.

[4] S. Berinato. The future of security. ComputerWorld, `http://www.computerworld.com/printthis/2003/0,4814,88646,00.html`, December 2003.

[5] D. D. Clark, C. Partridge, J. C. Ramming, and J. T. Wroclawski. A Knowledge Plane for the Internet. In *ACM SIGCOMM*, 2003.

[6] J. Considine, F. Li, G. Kollios, and J. W. Byers. Approximate aggregation techniques for sensor databases. In *ICDE*, 2004.

[7] A. Deshpande, S. K. Nath, P. B. Gibbons, and S. Seshan. Irisnet: Internet-scale resource-intensive sensor services. In *SIGMOD*, 2003.

[8] DShield: Distributed Intrusion Detection System. `http://www.dshield.org/`, June 2004.

[9] M. J. Freedman, E. Freudenthal, and D. Mazières. Democratizing content publication with coral. In *USENIX/ACM NSDI*, 2004

[10] M. Garofalakis and P. Gibbons. Approximate query processing: Taming the terabytes. In *VLDB*, 2001. Tutorial.

[11] S. D. Gribble, A. Y. Halevy, Z. G. Ives, M. Rodrig, and D. Suciu. What can database do for peer-to-peer? In *WebDB*, 2001.

[12] M. Gritter and D. R. Cheriton. An architecture for content routing support in the internet. In *USITS*, 2001.

[13] J. M. Hellerstein. Toward network data independence. *SIGMOD Record*, 32(3):34–40, 2003.

[14] R. Huebsch, J. M. Hellerstein, N. Lanham, B. T. Loo, S. Shenker, and I. Stoica. Querying the Internet with PIER. In *VLDB*, 2003.

[15] A. Jain, J. M. Hellerstein, S. Ratnasamy, and D. Wetherall. A wakeup call for internet monitoring systems: The case for distributed triggers. In *HotNets-III*, Nov. 2004.

[16] B. T. Loo, J. M. Hellerstein, and I. Stoica. Customizable routing with declarative queries. In *HotNets-III*, Nov. 2004.

[17] S. Nath, P. B. Gibbons, Z. Anderson, and S. Se. Synopsis diffusion for robust aggregation in sensor networks. In *SenSys*, 2004.

[18] J. C. Navas and M. J. Wynblatt. The network is the database: Data management for highly distributed systems. In *SIGMOD*, 2001.

[19] V. S. Pai, L. Wang, K. Park, R. Pang, and L. Peterson. The dark side of the web: An open proxy's view. In *HotNets-II*, Nov. 2003.

[20] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson. Characteristics of Internet background radiation. In *IMC*, 2004.

[21] M. A. Paskin and C. E. Guestrin. Robust probabilistic inference in distributed systems. In *UAI*, 2004.

[22] L. Peterson, T. Anderson, D. Culler, and T. Roscoe. A Blueprint for Introducing Disruptive Technology into the Internet. In *HotNets-I*, 2002.

[23] S. Shenker. The data-centric revolution in networking. In *VLDB*, 2003. Keynote talk. `http://pier.cs.berkeley.edu/shenker-vldb.ppt`.

[24] C. R. Simpson Jr. and G. F. Riley. NETI@Home: A Distributed Approach to Collecting End-to-End Network Performance Measurements. In *Passive and Active Measurement Workshop (PAM)*, 2004.

[25] I. Stoica, D. Adkins, S. Zhuang, S. Shenker, and S. Surana. Internet indirection infrastructure. In *ACM SIGCOMM*, 2002.

[26] The DIMES Project. `http://www.netdimes.org/`, June 2004.

[27] A. Vahdat, K. Yocum, K. Walsh, P. Mahadevan, D. Kostic, J. Chase, , and D. Becker. Scalability and accuracy in a large-scale network emulator. In *OSDI*, 2002.

[28] R. van Renesse, K. P. Birman, D. Dumitriu, and W. Vogel. Scalable management and data mining using astrolabe. In *IPTPS*, 2002.

[29] D. Wallach. A Survey of Peer-to-Peer Security Issues. In *Intl. Symp. on Software Security*, 2002.

[30] B. White, J. Lepreau, L. Stoller, R. Ricci, S. Guruprasad, M. Newbold, M. Hibler, C. Barb, and A. Joglekar. An integrated experimental environment for distributed systems and networks. In *OSDI*, 2002.

[31] X. Yang. Nira: a new internet routing architecture. In *SIGCOMM Workshop on Future Directions in Network Architecture (FDNA)*, 2003.

[32] V. Yegneswaran, P. Barford, and J. Ullrich. Internet intrusions: Global characteristics and prevalence. In *SIGMETRICS*, 2003.

[33] M. Zhang, C. Zhang, V. Pai, L. Peterson, and R. Wang. Planetseer: Internet path failure monitoring and characterization in wide-area services. In *OSDI*, 2004.